DOCUMENT RESUME

ED 228 289 TM 830 197

AUTHOR Plake, Barbara S.; And Others
TITLE Differential Performance of Males and Females on Easy to Hard Item Arrangements; Influence of Feedback at the Item Level.
PUB DATE Apr 83
NOTE 18p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, April 12-14, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Difficulty Level; *Feedback; Higher Education; Scoring; *Sex Differences; Test Bias; *Test Format; *Testing Problems; *Test Items
IDENTIFIERS Answer Until Correct; Differential Reinforcement; *Item Position (Tests); Latin Squares

ABSTRACT

Differential test performance by undergraduate males and females enrolled in a developmental educational psychology course (n=167) was reported on a quantitative examination as a function of item arrangement. Males were expected to perform better than females on tests whose items arranged easy to hard. Plake and Ansorge (1982) speculated this may be a function of differential reinforcement backgrounds of males and females. This study used a Latin Square design and provided item performance feedback for a nonquantitative examination. A scoring method was used which allowed students to ascertain, on each item as they proceeded through the test, whether their answer was correct or not. To control for effects of test specificity the study involved three course examinations given throughout the semester. Significant sex-by-order effects did not occur but the lack of this effect could possibly be due to the use of a nonquantitative examination. It still remains to be seen whether, in using a quantitative examination, differential effects of item feedback may be a source of explanation for the evidence for sex-by-order effects. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED228289

Differential Performance of Males and Females on Easy to Hard Item Arrangements: Influence of Feedback at the Item Level

Barbara S. Plake
University of Nebraska-Lincoln
Gerald J. Melican
Educational Testing Service
Linda Carter and Mike Shaughnessy
University of Nebraska-Lincoln

Paper Presented at the Annual Meeting of the National Council in Education, Montreal, Canada, April, 1983

Running Head: Differential Performance

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

B S. Plake

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

PRINTED IN U.S.A

TM 830 157

Differential Performance of Males and Females on Easy to Hard Item Arrangements: Influence of Feedback at the Item Level

Barbara S. Plake
University of Nebraska-Lincoln
Gerald J. Melican
Educational Testing Service
Linda Carter and Mike Shaughnessy
University of Nebraska-Lincoln

Abstract

Differential test performance by males and females was reported on a quantitative examination as a function of item arrangement (Plake, Ansorge, Parker,and Lowry, 1982) whereby males performed better than females on tests whose items arranged easy to hard. Plake and Ansorge, 1982, speculated this may be a function of differential reinforcement backgrounds of males and females. This study used a Latin Square design and provided item performance feedback for a nonquantitative examination. Significant sex by order effects did not occur but the lack of this effect could possibly be due to the use of a nonquantitative examination (Plake and Ansorge, in press). It still remains to be seen whether, in using a quantitative examination, differential effects of item feedback may be a source of explanation for the evidence for sex by order effects.

Differential Performance of Males and Females on Easy to Hard Item Arrangements: Influence of Feedback at the Item Level

In their study of the effects of item arrangements on test performance, Plake, Ansorge, Parker, and Lowry (1982) found a differential effect of item arrangement for a mathematics examination for males and females. The male subjects performed systematically superiorly to females when the items were in an easy to hard arrangement. This arrangement produced the highest scores for the males, across the arrangements investigated (easy to hard, spiral cyclical, and random) and the lowest for the females. Plake and Ansorge (1982) attempted to replicate this finding using a non-mathematical content area. Although their results were not statistically significant due in part to limited sample size, trends in the data suggested a consistent pattern of performance for males and females across the item arrangements. Male scores, for the non-mathematical content area, were highest again for the easy to hard arrangement and exceeded the female performance. As was found with the mathematics content area, the females' scores were lowest on the test when the items were arranged in an easy to hard fashion.

In their 1982 study, Plake and Ansorge speculated on the reasons why male and females show differential sensitivity to item arrangement, particularly the easy to hard arrangement. They suggest that if success on an item in a test is viewed as a form

of positive reinforcement, and failure to answer an item correctly a form of negative reinforcement, it might be possible to hypothesize that the difference in item performance between males and females may be a function of differential reinforcement histories or backgrounds between males and females. When items are arranged from easy to hard, the examination questions become increasingly difficult. It is possible that the male examinees are more persistent or challenged to continue a sincere testing effort as the positive reinforcement pattern diminishes, whereas female examinees are more susceptive to extinction of effort as the positive reinforcement decreases and negative reinforcement increases. However, since within their study, examinees may have suspected but did not actually know if they got an item right or wrong on the test, the degree of positive and/or negative reinforcement may have been subtle at best.

The purpose of this study was to further investigate the effect of differential item performance by males and females on tests which have different item arrangements. A scoring method was used in the study which allowed students to ascertain, on each item as they proceeded through the test, whether their answer was correct or not. Therefore, this study allowed for a more sensitive evaluation of whether differential sensitivity to reinforcement strategies is a factor in performance discrepancies for males and females. In addition, to control for effects of test specificity the study involved three course examinations given throughout the semester of an undergraduate course in

developmental psychology. A Latin Square design was utilized so that every student received tests involving easy to hard, spiral cyclical and random arrangements in the three testing periods in the semester. Further, for some students, their first examination was arranged in an easy to hard fashion, their second examination had items which followed a spiral cyclical arrangement and their third one had items that were in a random order. Another group of students found their three examinations arranged in random, then easy to hard, and finally spiral cyclical arrangements. For the third group of students, the first examination was arranged in a spiral cyclical order, second one was in a random arrangement, and the third examination was easy to hard. Therefore, each item arrangement was in the first, second, and third testing position.

Procedure

The subjects were 167 students (39 males and 128 females) in an undergraduate course in developmental educational psychology at a large midwestern University, spring, 1982. The items for the three course examinations were selected from a pool of items which had been piloted on previous classes. For each of the three examinations, three forms of tests were prepared: (1) easy to hard, (2) spirical cyclical, and (3) random. Students within the class were randomly assigned in to three groups which then received their three exams: (1) easy to hard, spiral cyclical, random, (2) random, easy to hard, spiral cyclical, (3) spiral cyclical, random, easy to hard. The examinations were prepared so

that answer response cards were used by the students when taking the examinations. Each examination consisted of 20 questions. These cards allowed the student to answer each item until their answer was correct; the scoring system penalized to student for incorrect attempts to answer each question. The examinations were used, along with the final and a class assignment, in determining course grades.

Results

Means, standard deviations, and cell sizes for the raw data are displayed in Table 1. The data were subjected to a 3x3x3x2 Latin Squares least squares analysis of variance (Winer, 1971) with test, arrangement, group and sex as independent variables and number right as the dependent variable. This analysis allows for the investigation of test, arrangement, group, and sex main effects and double interactions of test x sex, arrangement x sex and group x sex. No other interactions are assessible.

A significant sex effect was found ($F(1,483)=53.76, p<.01$) where females, across the three test forms, scored significantly higher than their male counterparts. No other effects were statistically significant (=.05). These results of the summary table are shown in Table 2.

In particular, no sex by arrangement effect was found. To further investigate the impact of sex x arrangement on a nonquantitative examination, the mean scores across these test

arrangements, regardless of sequence, were determined. These values are shown in under the Total heading in Table 2 and displayed in Figure 1. Upon inspecting these values, it is evident that something quite different is happening on the easy to hard arrangements for male and female subjects than was reported previously by Plake et al. (1982). Both male and female subjects perform least well on the easy to hard arrangement.

Although group x sex was not significant at =.05, it would have been significant if was set at .10. Even though this result was not one of fundamental interest to the study, the influence of the sequence of tests shows a dramatically non parallel pattern for male and female subjects. This result is displayed in Figure 2 and the means by sequence for males and females are shown in Table 3. The favored sequence for males is obviously the spiral cyclical,random, easy to hard sequence while no meaningful differences in performance for males is revealed when the average performance across the easy to hard, spiral cyclical, random sequence is compared to that for the random, easy to hard, spiral cyclical sequence. When considering the trends in performance for females subjects, their highest overall performance was obtained for test sequences of easy to hard, spiral cyclical, random, second best for random, easy to hard, spiral cyclical and least well for spiral cyclical, random, easy to hard sequences. None of the pairwise comparisons of performance by the males was significant (=.05), while all pairwise comparisons for females were. This fact is due in part to the differential sample sizes of males and females.

Conclusions

The expected superiority of male subjects on tests with items arranged in an easy to hard fashion did not happen. In contrast, males performed, in general, least well when their test were arranged in an easy to hard fashion.

These results are perplexing and need further clarification. A recent study by Plake and Ansorge (in press) which failed to find parallel effects between quantitative and nonquantitative examinations with respect to item arrangement and sex on the subject on cognitive and perception scores brings into question the comparability of item arrangement effects across content areas.

This research did not answer the question it set out to address: why males perform systematically better on tests whose items are arranged in an easy to hard format. It appears from this work and others (Plake and Ansorge, in press) that test content area is a crucial variable which may effect when and if sex differences arise in test performance on tests with items arranged easy to hard. Interpretation is further complicated by whether the mean difficulty level of the test in general will influence how students will perform on the various test item arrangements. For example, for this study, Test 1 had a mean item difficulty level of .32 (sd=.08), for Test 2, the mean item difficulty level was .48 (sd=.11), while for Test 3, the mean item

difficluty value was .34 (sd=.12). The quantitative test for which a significant sex by arrangement effect, favoring males was revealed, Plake et al.,1982, had mean item difficulty equaling .55 (sd=.11).

Further research, then, needs to focus on the following pertinent issues:

(1) Is the sex by arrangement effect favoring males on easy to hard arrangements repeatable in a quantitative examination?

(2) If a repeatable sex by arrangement effect favoring males on easy to hard arrangements exists, can this effect be better understood by looking into reinforcement theory by providing feedback at the item level?

(3) Are there consistent non parallel arrangement effects between quantitative and nonquantitative examinations?

(4) What are the dimensions of the effects of item arrangements on nonquantitative examinations? Does the sequence of item arrangements across tests make a difference in performance by males and females? How do sequence effects influence item arrangement effects on quantitative exams? Are the sequence effects (if any) comparable for quantitative and nonquantitative exams?

(5) What is the relationship between overall test item difficulty and the existence of a sex x arrangement effect in general? Do these effects vary between quantitative and nonquantitative content areas?

This study, therefore, may be of more value than its results

indicate because of the nature of the questions that were generated based on the findings reported here.

References

Plake, B. S. and Ansorge, C. J. Effects of Item Arrangement, Sex of the Subject, and Test Anxiety on Cognitive on Perception Scores on an Educational Psychology Examination, National Council on Measurement in Education, New York, 1982.

Plake, B. S., Ansorge, C. J. Effects of item arrangement, sex of the subject and test anxiety on cognitive and perception scores in a nonquantitative content area. Journal of Experimental Education, in press.

Plake, B. S., Ansorge, C. J., Parker, C. S., and Lowry, S. R., Effects of item arrangement, knowledge of arrangement, test anxiety, and sex on test performance. Journal of Educational Measurement, 1982, 19, (1), 49-57.

Winer, B. J. Statistical Principles in Experimental Design, Second ed., New York: McGraw Hill, 1971, p. 702-703.

Acknowledgements

The authors thank Barbara Carroll for her assistance with the data analysis.

Table 1

Means, Standard Deviations and Sample Sizes

Arrangement			Test Time 1	2	3	Total
Easy to Hard	Males	N	12	15	12	39
		$\bar{X}$	44.50	42.53	44.50	43.74
		s	8.06	6.03	8.33	7.30
	Females	N	41	46	41	128
		$\bar{X}$	49.85	49.41	48.02	49.11
		s	5.89	5.29	6.77	5.99
Spiral Cyclical	Males	N	12	12	15	39
		$\bar{X}$	46.33	45.75	43.73	45.15
		s	5.84	2.67	7.44	5.80
	Females	N	41	41	46	128
		$\bar{X}$	48.71	50.00	49.24	49.31
		s	5.41	4.74	6.11	5.46
Random	Males	N	15	12	12	39
		$\bar{X}$	45.47	47.33	41.75	44.90
		s	4.44	6.41	5.75	5.82
	Females	N	46	41	41	128
		$\bar{X}$	48.09	48.90	50.34	49.07
		s	5.37	6.39	5.74	5.86

Table 2

ANOVA Summary Table

Least Squares Analysis

Source	df	SS	MS	F
Order	2	19.64	9.82	0.28
Time	2	26.94	13.47	0.39
Group	2	89.75	44.87	1.29
Sex	1	1869.43	1869.43	53.76**
Order x Sex	2	28.70	14.35	0.41
Time x Sex	2	88.88	44.44	1.28
Group x Sex	2	176.56	88.28	2.54
Within	483	16796.82	34.78	

** $p<.01$

Table 3

Test Performance by Males and Females

for Test Item Arrangement by Test Arrangement Sequence

	Test			
Males	Easy to Hard	Spiral-Cyclical	Random	Total
EH-SC-R	44.50	45.75	41.75	44.00
R-EH-SC	42.53	43.73	45.47	43.91
SC-R-EH	44.50	46.33	47.33	46.06
Females				
EH-SC-R	49.85	50.00	50.34	50.07
R-EH-SC	49.41	49.24	48.09	48.91
SC-R-EH	48.02	48.71	48.90	48.54

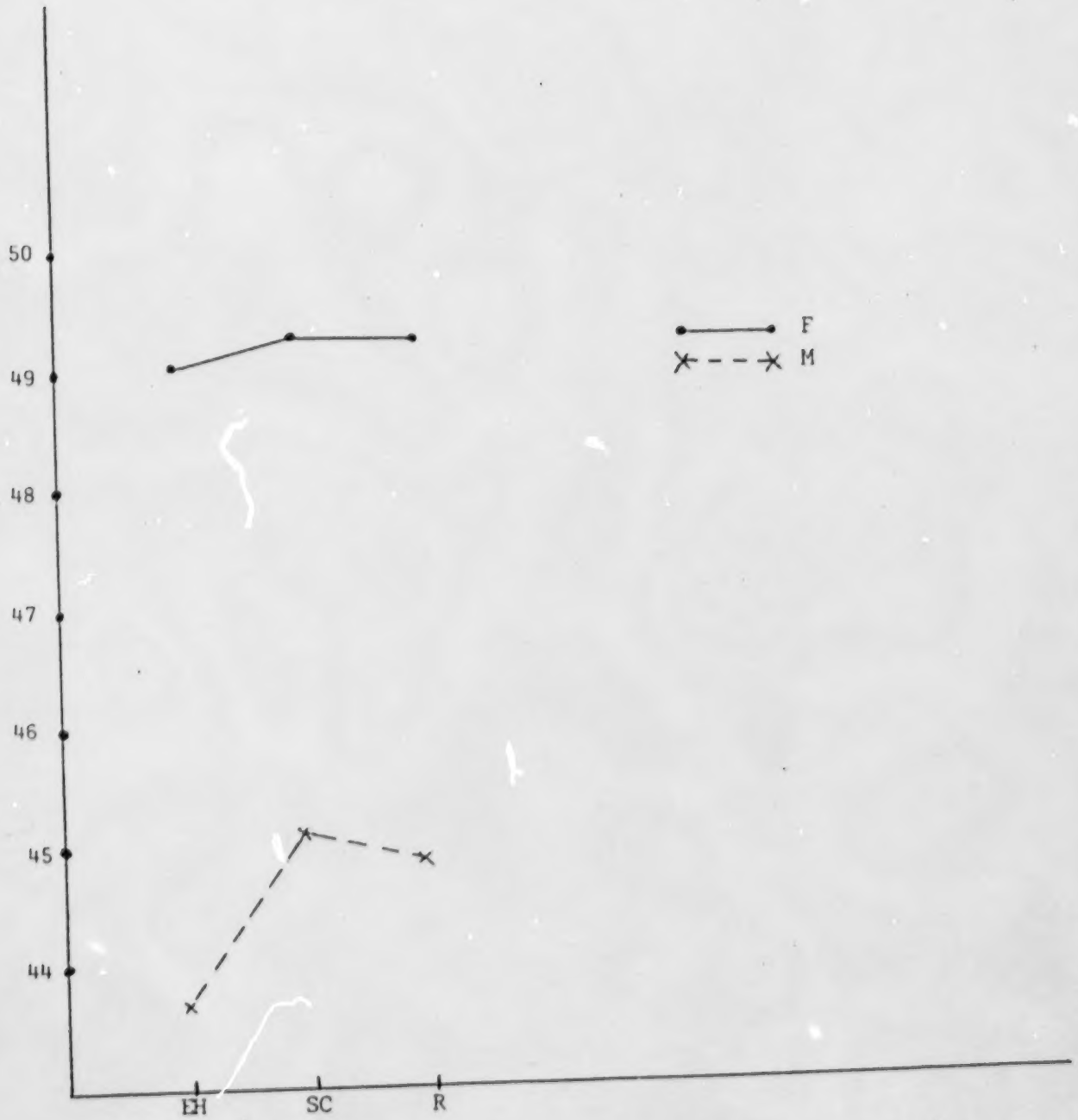

Figure 1. Mean test score across test sequences for EH,SC, and R arrangements for male and female subjects

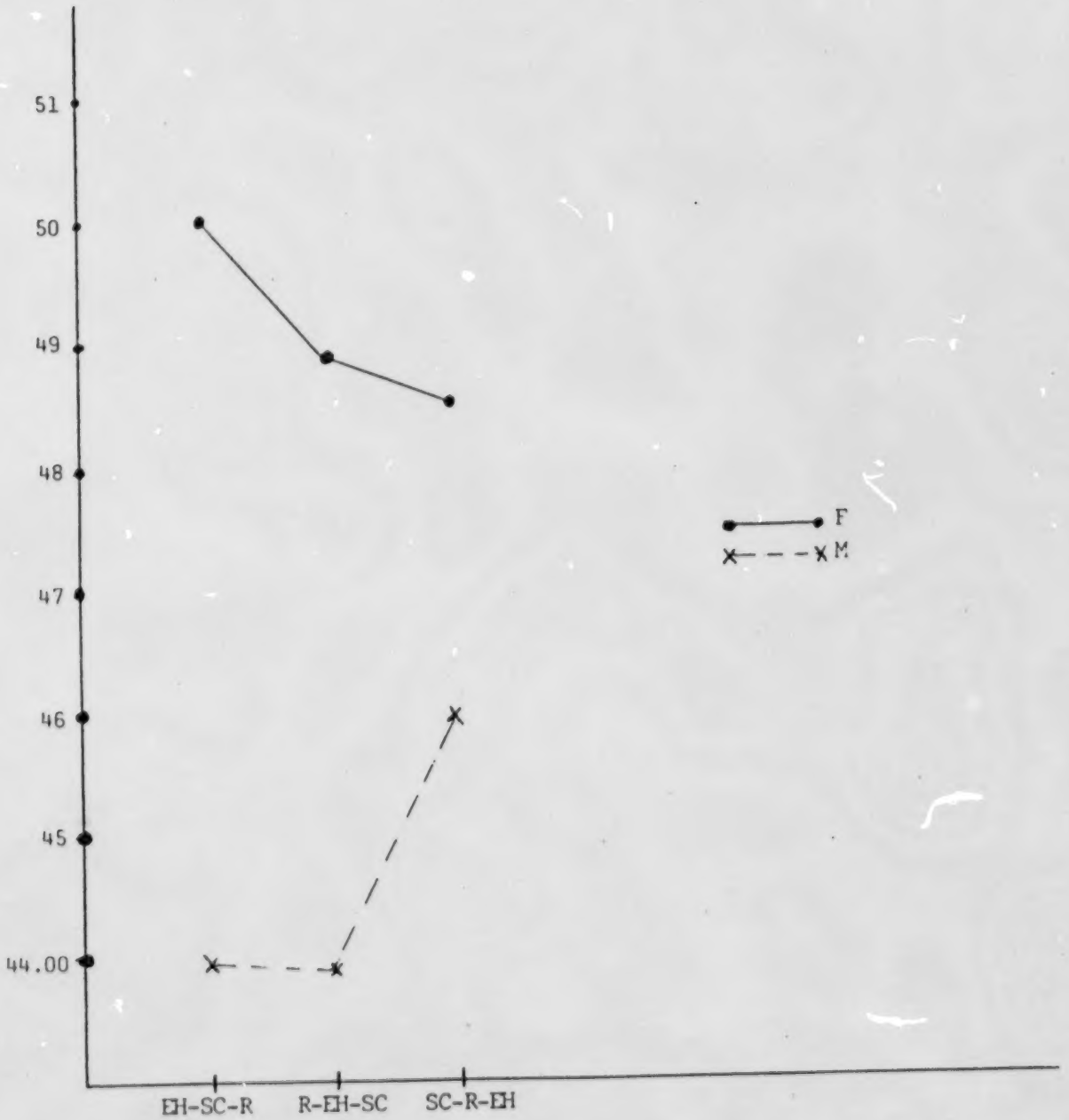

Figure 2: Average performance across item arrangements for test sequence by sex